week11-twitter

Mathi Manavalan

3/31/2020

R Studio API Code

Setting the working directory is not necessary for an RMarkdown file.

Libraries

Importing the libraries necessary for the data importing, cleaning, and analyzing.

```
library(twitteR)
library(tidyverse)
```

Data Import and Cleaning

The keys inputted into the function below are the keys found for my particular app through my Twitter developer account.

```
setup_twitter_oauth(api, secretKey, token, secretToken)
```

```
## [1] "Using direct authentication"
```

```
tweets <- searchTwitter("#rstats", 1000)
tweets_clean <- strip_retweets(tweets)
tweets_tbl <- twListToDF(tweets_clean) %>%
  select(screenName, text, favoriteCount, retweetCount)
```

Analysis

```
plot1sum <- summary(lm(tweets_tbl$retweetCount ~ str_length(tweets_tbl$text)))
plot1sum
```

```
##
## Call:
## lm(formula = tweets_tbl$retweetCount ~ str_length(tweets_tbl$text))
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -6.416 -5.049 -2.318  0.157 45.976
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -4.71988    3.28252  -1.438  0.15341
## str_length(tweets_tbl$text)  0.07733    0.02831   2.732  0.00739 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.778 on 106 degrees of freedom
## Multiple R-squared:  0.06576,    Adjusted R-squared:  0.05695
## F-statistic: 7.461 on 1 and 106 DF,  p-value: 0.007387
```

The above statistical test creates a linear model between the number of characters in a tweet and how many times that tweet was retweeted.

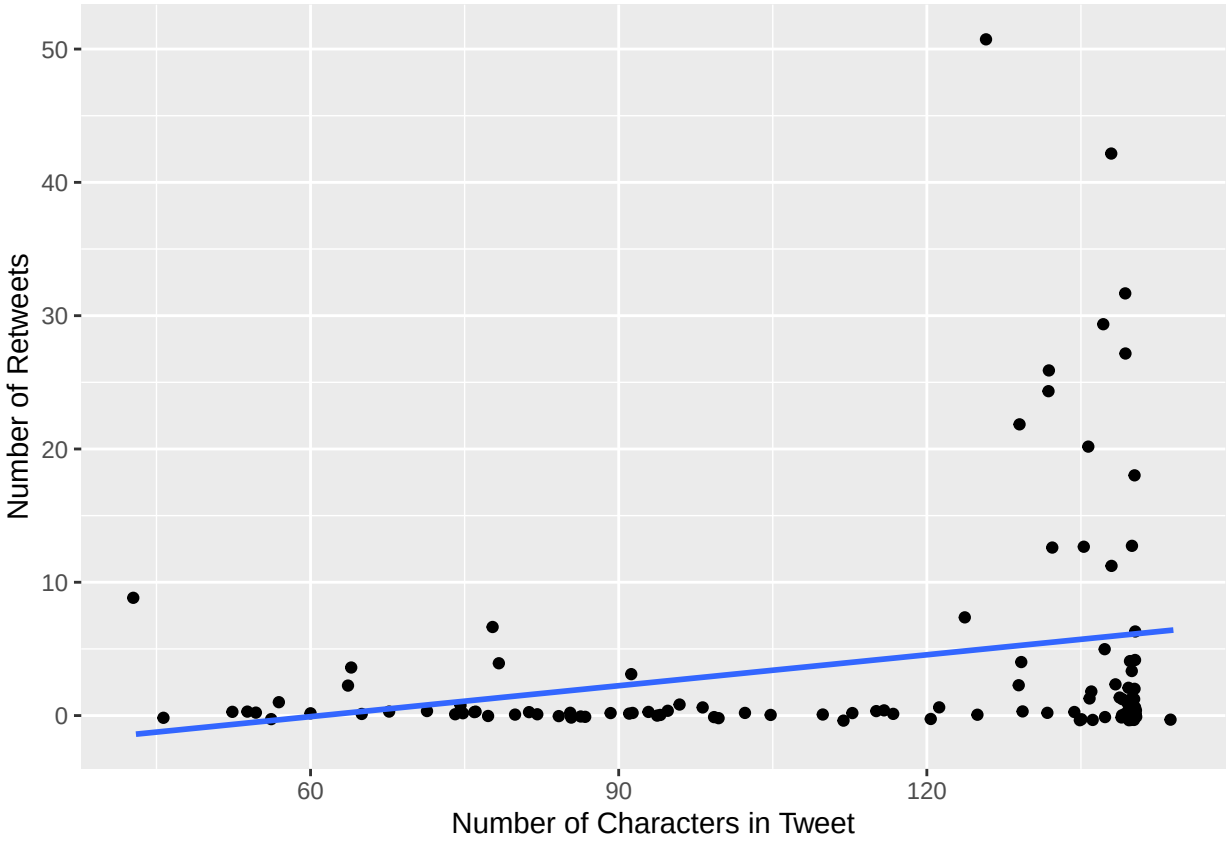
```
plot2sum <- summary(lm(tweets_tbl$favoriteCount ~ str_length(tweets_tbl$text)))
plot2sum
```

```
##
## Call:
## lm(formula = tweets_tbl$favoriteCount ~ str_length(tweets_tbl$text))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.489 -5.067 -3.631 -0.409 85.511
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.51630    4.38404   0.118    0.906
## str_length(tweets_tbl$text)  0.04266    0.03781   1.128    0.262
##
## Residual standard error: 11.72 on 106 degrees of freedom
## Multiple R-squared:  0.01187,    Adjusted R-squared:  0.002546
## F-statistic: 1.273 on 1 and 106 DF,  p-value: 0.2617
```

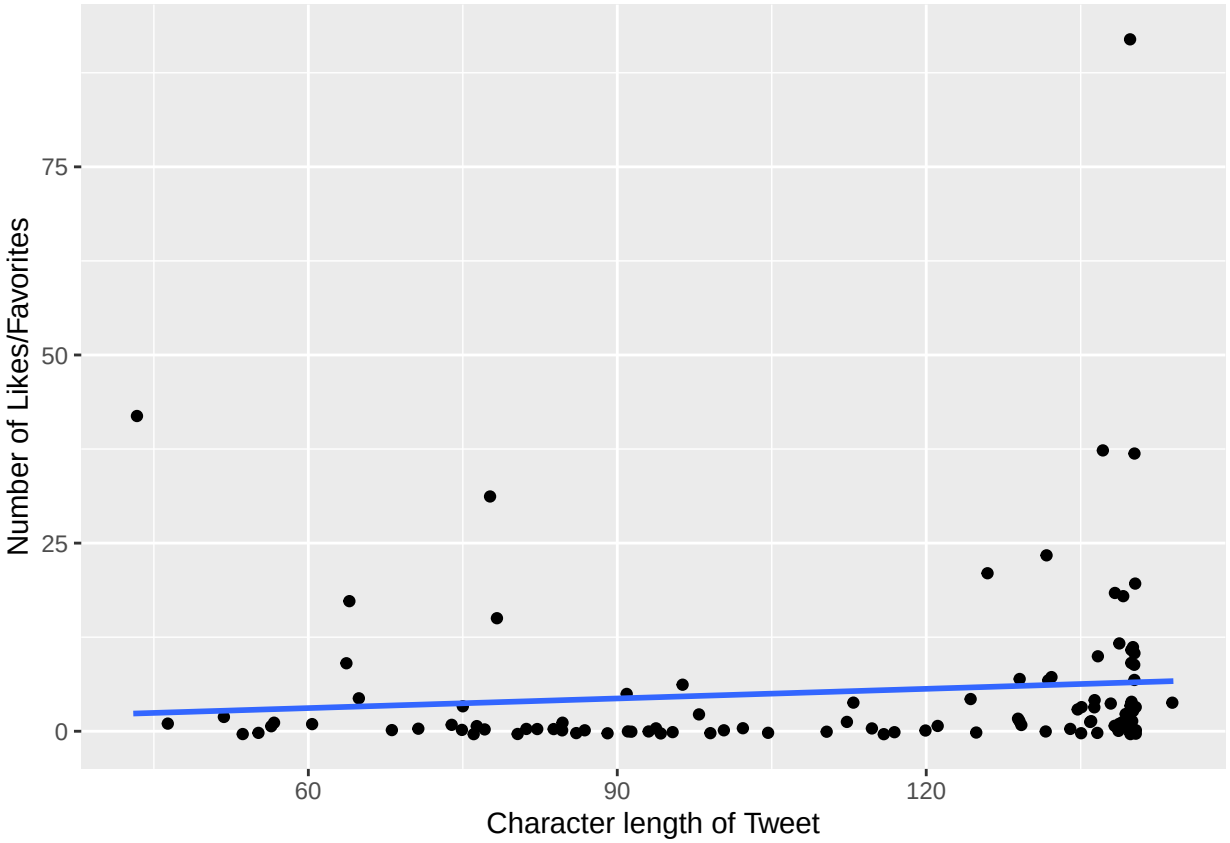
The above statistical test creates a linear model between the number of characters in a tweet and how many times that tweet was liked/favorited.

Visualization

```
plot1 <- ggplot(tweets_tbl, aes(x=str_length(text), y= retweetCount)) +
  geom_jitter() +
  labs(x = "Number of Characters in Tweet", y = "Number of Retweets") +
  geom_smooth(method = "lm", se = FALSE)
plot1
```



The above plot displays the relationship between the number of characters in a tweet (on the x-axis) and how many times that tweet was retweeted (on the y-axis) as a scatterplot. The plot also includes a linear regression line between the two variables.



The above plot displays the relationship between the number of characters in a tweet (on the x-axis) and how many times that tweet was 'liked' or 'favorited' (on the y-axis) as a scatterplot. The plot also includes a linear regression line between the two variables.